

Identifying User Traits by Mining Smart Phone Accelerometer Data

Gary M. Weiss and Jeffrey W. Lockhart
Department of Computer and Information Science
Fordham University
441 East Fordham Road
Bronx NY 10458

{gweiss, lockhart}@cis.fordham.edu

ABSTRACT

Smart phones are quite sophisticated and increasingly incorporate diverse and powerful sensors. One such sensor is the tri-axial accelerometer, which measures acceleration in all three spatial dimensions. The accelerometer was initially included for screen rotation and advanced game play, but can support other applications. In prior work we showed how the accelerometer could be used to identify and/or authenticate a smart phone user [11]. In this paper we extend that prior work to identify user *traits* such as sex, height, and weight, by building predictive models from labeled accelerometer data using supervised learning methods. The identification of such traits is often referred to as “soft biometrics” because these traits are not sufficiently distinctive or invariant to uniquely identify an individual—but they can be used in conjunction with other information for identification purposes. While our work can be used for biometric identification, our primary goal is to learn as much as possible about the smart phone user. This mined knowledge can be then be used for a number of purposes, such as marketing or making an application more intelligent (e.g., a fitness app could consider a user’s weight when calculating calories burned).

Categories and Subject Descriptors

H.2.8 [Database Applications]: Data Mining

General Terms

Algorithms, Measurement, Experimentation, Human Factors.

Keywords

Sensor mining, sensors, biometrics, cell phones, smart phones, data mining, accelerometer.

1. INTRODUCTION

Smart phones are quite sophisticated and increasingly incorporate diverse and powerful sensors. These devices provide unprecedented opportunities for sensor mining since they include a large variety of sensors, including an: acceleration sensor (accelerometer), location sensor (GPS), direction sensor (compass), audio sensor (microphone), image sensor (camera), proximity sensor,

light sensor, and temperature sensor. Because of their portability and ubiquity, these smart phone sensors provide us with an opportunity to learn a great deal about their users. In prior work, for example, we used the smart phone’s accelerometer to determine *what* physical activity (e.g., walking, jogging, sitting) a user is performing [10] and to identify/authenticate the user [11]. In this paper we extend this latter work on biometric identification to predict user characteristics, or *traits*.

The identification of personal traits is often referred to as soft biometrics because these traits, on their own, are not distinctive enough, or invariant enough, to uniquely identify an individual [9]. Soft biometric traits include the three traits that we predict in this paper—sex, height, and weight—which clearly are less distinctive or invariant than hard biometric traits like fingerprints. Nonetheless, these “soft” traits can be used in conjunction with other information or other traits to improve the accuracy of a biometric identification system [2,9,14]. While the work described in this paper can be used for biometric identification, our goal is far more general—we want to learn as much as possible about people (in this case smart phone users). This mined knowledge can be used for a variety of purposes. For example, this knowledge can be used for marketing and advertising, since most soft traits, like sex, height, weight, hair color, physical activity level, and foot/shoe size could have an impact on what content advertisers might want to present. Knowledge of these soft traits can also be used to make a smart phone application behave more intelligently by having it automatically customize its behavior to the user. For example, a fitness or diet “app” could consider your weight when calculating calories burned. There are undoubtedly many additional uses for this type of information, although clearly privacy and security issues should be considered.

The work in this paper is part of a larger effort by the WISDM (Wireless Sensor Data Mining) research group [21] to develop smart phone-based sensor mining applications. The applications are built on the hardware and software platform that we are developing [12]. Currently, this platform relies on Android-based smart phones, which were used for all of the experiments described in this paper. All of our work thus far (activity recognition, biometric identification, and now trait prediction) relies only on the tri-axial accelerometer that is included in all Android smart phones. This sensor, which was originally included mainly for screen rotation and advanced game play, measures acceleration in all three spatial dimensions. In the future, we plan to investigate the use of other sensors for sensor mining applications, including trait prediction. All of the predictive models built in this paper were generated from the accelerometer data by using standard predictive data mining methods.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SensorKDD '11, August 21, 2011, San Diego, CA, USA.

Copyright 2011 ACM 978-1-4503-0832-8...\$10.00.

2. DATA COLLECTION

In this section we describe our process for collecting labeled accelerometer data, which is used for generating and evaluating the trait-prediction models. Our data was collected from volunteers who carried one of the Android smart phones provided by our team. The volunteers were asked to walk, with the Android phone in their pocket, for approximately 5 to 10 minutes. Our data is based on 70 users, but because we had previously collected data for our activity recognition research [10], many of the users had data for additional activities besides walking; for this study the data from these extra activities was ignored. This data collection effort was approved by Fordham’s Institutional Review Board, since we were “experimenting” on human subjects and there was some slight risk of injury (e.g., the subject could fall). All subjects provided written informed consent prior to participating in our study and were also asked to fill out a detailed questionnaire. This questionnaire asked the user for information about all soft traits that we could think of that might possibly be predicted from a smart phone’s accelerometer data—including the three traits studied in this paper. Participants were permitted to skip any questions they wanted to, without being dismissed from the study. As a consequence, there were fewer than 70 participants for the sex, height and weight prediction tasks. The details associated with the generated data sets are provided in Section 3.3.

The soft traits that appeared on our questionnaire are listed in Table 1. This paper does not analyze most of these traits. One reason is due to a lack of training data. Because we collected data from only 70 participants, we could not realistically analyze those traits that permitted a great variety of class values or were highly imbalanced. For example, we could not properly predict race based on the accelerometer data because certain racial groups had very little representation. Given that the time required to collect data for a 5 minute walk is far less than the time required to collect data for our activity recognition task (5 minutes vs. 60 minutes), we do expect to radically increase the size of our data set in the near future. But we did run experiments on some of these other traits and the results were not very good. We simply did not have sufficient time to study these prediction tasks, but plan to return to them shortly. We are hopeful that with additional training data and by constructing better features (Section 3.2), we will be able to predict some of these other traits. However, it is very possible that some of these traits may simply not be predictable from 5 or 10 minutes of accelerometer data.

One issue with our data collection is that we selected our data from users who were, in many ways, very homogenous. For example, since almost all study participants were solicited on our campus, almost all were college students. As a consequence, we could not meaningfully predict age, an important trait that may be predictable, because most of our users were between the ages of 18 and 24. In the future we do plan to draw more volunteers from alternate locations and to target users with traits that are underrepresented in our data set.

Several of the soft traits listed in Table 1 were included in our survey specifically because we believed that they may impact gait and thus impact the accelerometer data. For example, if one normally carries a backpack it could impact one’s gait and, similarly, the number of hours of aerobic exercise could also impact gait. Note that the last three “traits” listed in Table 1 are not what would normally be considered user traits (e.g., type of footwear), but rather are things associated with the user during the data collection process, which may impact the user’s gait. We collect this

information so we can analyze their impact on the accelerometer data and, more importantly, so that we can try to predict them. The traits listed in Table 1 were selected based on the assumption that accelerometer data was being collected. If we were collecting data from other sensors we certainly would have come up with a different list (e.g., if audio data was collected from the phone’s microphone we might ask about the volume of the user’s normal speaking voice). We do hope to extend the work in this paper to other sensors in the future.

Table 1. Soft Traits that were Collected

Traits that we currently analyze

- Sex: biological sex {male, female}
- Height: measured in inches
- Weight: measured in pounds

Traits that we do not currently analyze

- Age: measured in years (*mainly college students*)
 - Race: {White, Black, Asian/Pacific Islander, Hispanic or Latino, Other}
 - Area you grew up in {Rural, Urban, Suburban}
 - Handedness: {Right, Left, Ambidextrous}
 - Hours of aerobic exercise per week
 - Type of bag do you carry {Backpack, Briefcase, ...}
 - Play organized sports in last 12 months {Yes, No}
 - Do you have an injury that affects the way you walk: {Yes, No}
 - Shoe Size: also specifies if men’s or women’s size
 - Hours of academic work per week excluding lecture
 - Footwear during walking {Sneakers, High Tops, ...}
 - Clothing during walking {Shorts, Pants, Dress, Other}
 - Phone Position during walking (Belt, Left Pocket, Right Pocket)
-

Data collection was controlled by our WISDM Sensor Collector application, which can be downloaded for free from the Android marketplace. This application, through a simple graphical user interface, permits us to record the user’s name, to start and stop data collection, and label the activity being performed (in this case walking). Through the application we can control what sensors are being monitored as well as how frequently the sensor data is collected. In all cases we collected the accelerometer data using our default sampling frequency of 50ms, yielding 20 samples per second. Data collection was supervised by one of the WISDM team members to ensure the quality of the data (and that no one ran off with our \$500 phones). Data was then collected either directly from the phone via a USB connection or transmitted over a cellular connection to our Internet-connected WISDM server.

3. DESCRIPTION OF EXPERIMENTS

Our experiments are described in this section and then the results from these experiments are presented and analyzed in Section 4. In Section 3.1 we define the prediction tasks associated with predicting the three user traits. Then in Section 3.2 we discuss the process for transforming the raw time-series accelerometer data

into examples so that traditional data mining prediction algorithms can be used. This step mainly involves the construction of higher level features. Section 3.3 describes the “transformed” data sets generated by this transformation step, where each data set is associated with one of the prediction tasks described in Section 3.1. Finally, in Section 3.4 we describe the prediction algorithms and the methodology used to build and evaluate the predictive models. The data transformation step described in Section 3.2 is essentially identical to the one used in our prior work [10-12] but the other sections are all new.

3.1 User Trait Prediction Tasks

We cover three basic trait prediction tasks in this paper: sex prediction, height prediction, and weight prediction. The first task, sex prediction, is fundamentally a binary classification task while the other two tasks are fundamentally regression (i.e., numerical prediction) tasks. Specifically, sex prediction involves classifying a smart phone user as either male or female while the height and weight prediction tasks involve predicting the user’s height in inches or weight in pounds. The class labels for these tasks are acquired via the questionnaire described in Section 2.

The interpretation and evaluation of numerical predictions is not as straightforward as for classification problems. While there are statistics like root mean squared error for characterizing the prediction performance, it is still often difficult to assess the quality of the results. Thus, in addition to treating the height and weight prediction tasks as regression tasks, we also convert them into classification tasks; for some applications this formulation may actually be of more use. We use the mappings shown in Table 2 to convert the height and weight prediction tasks into classification tasks. The distribution of people between these classes is approximately equal, but since the measurements were all recorded using integers, the resulting class distribution is not perfectly balanced. The precise distribution information is provided in Section 3.3, which describes the data sets.

Table 2. Height and weight categories

Height (inches)			Weight (lbs.)		
From	To	Class	From	To	Class
0	65	Short	0	139	Light
66	69	Medium	140	169	Medium
70	∞	Tall	170	∞	Heavy

One issue for the two classification problems associated with Table 2 is that a classifier will have difficulty distinguishing examples at the borders between the classes. Given that at this stage in our research we are mainly concerned with ensuring that our models can identify substantial differences, our experiments in this paper exclude the center class (*medium* for both tasks) from the data set, so that *medium* people do not appear in either the training or testing sets. Thus, our task in these cases is to determine whether we can distinguish a *short* person from a *tall* person and a *light* person from a *heavy* person.

3.2 Data Transformation

We first transform the raw time series data into examples, since the classification algorithms that we use in this paper cannot directly learn from time series data [22]. To accomplish this we

divided the data into 10-second segments and then generated features from the accelerometer values contained in each 10-second interval. Since acceleration data is collected for 3 axes 20 times per second, each 10-second interval has 600 total values. We refer to this 10-second interval as the example duration (ED). We chose a 10-second example duration because we felt that it provided sufficient time to capture several repetitions of the motions involved in walking and because preliminary experiments indicate that this value provides good results for our current applications.

The data contained in one example duration is converted into a single example, described by forty-three features. These forty-three features are variations of just the six basic features, listed below (the number of features associated with each basic feature is noted in brackets):

- Average[3]: Average acceleration value (for each axis)
- Standard Deviation[3]: Standard deviation (for each axis)
- Average Absolute Difference[3]: Average absolute difference between the value of each of the 200 readings within the ED and the mean value over those 200 values (for each axis)
- Average Resultant Acceleration[1]: Average of the square roots of the sum of the values of each axis squared $\sqrt{(x_i^2 + y_i^2 + z_i^2)}$ over the ED
- Time Between Peaks[3]: Time in milliseconds between peaks in the sinusoidal waves associated with most activities (for each axis)
- Binned Distribution[30]: The values for each axis are divided into 10 equally spaced bins and then the fraction of the 200 values that fall within each of the bins is recorded.

The “time between peaks” feature requires further explanation. Walking and other repetitive activities tend to generate repeating waves for some or most of the axes, and this feature estimates the time between successive peaks. To estimate this value we find the highest peak within the record for each dimension, set a threshold based on a percentage of this value, and then find other peaks that meet or exceed this threshold; if no peaks meeting this criterion are found, then the threshold is lowered until we find at least three peaks. We then measure the time between successive peaks and calculate the average. For samples where at least three peaks could not be found, the distance between peaks is marked as unknown. This method was able to find the time between peaks for data that was collected while users walked. Certainly more sophisticated schemes will be tried in the future.

3.3 Description of Data Sets

All of the data sets employed in this study are created from an initial data set that contains data from 70 study participants. Many of these participants performed more activities than walking, due to our prior work on activity recognition [10], but for the purposes of this study, all data not associated with walking was removed. Because participants were not required to answer all of the survey questions, our task-related data sets do not contain data from all 70 participants. Overall, for the three traits that we study, the following numbers of users reported the necessary trait values: 66 for sex, 61 for height, and 63 for weight. The data set for sex is easiest to describe: it contains 38 (57.6%) male and 28 (42.4%) female participants.

The height data is relatively uncomplicated and, as shown in Figure 1a, the distribution is relatively symmetric. The average

height is 69.1 inches (5 foot 9 inches). The figure also shows the breakdown of users between the *short* and *tall* categories (*medium* users fall between these two categories). The *short* category contains 20 users, the *medium* category 19 users, and the *tall* category 22 users. For our binary classification problem that ignores the *medium* category, the class distribution is 47.6% *short* and 52.4% *tall*.

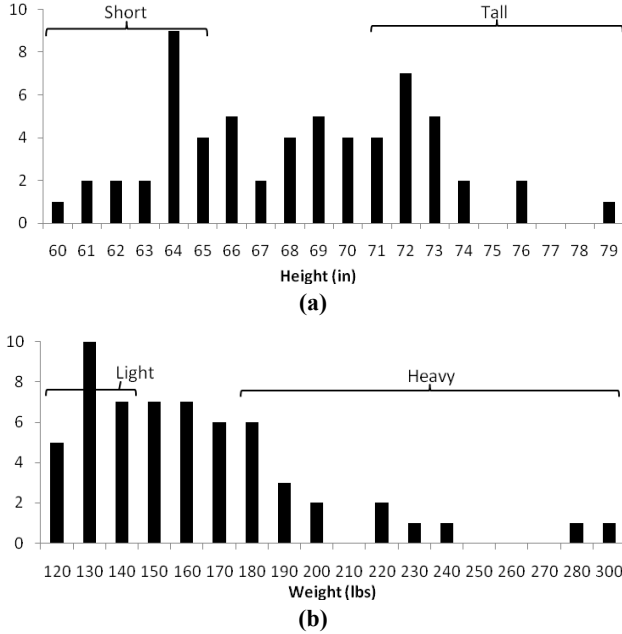


Figure 1. Distribution of (a) height and (b) weight values

The distribution of the weight data is presented in Figure 1b, which clearly shows an asymmetric distribution with a much longer tail toward the higher weights. Due to the larger range of numerical values than for height, the users are aggregated into 10 pound bins. What is not visible is that most users reported their weights in 5 pound increments. The 63 users that reported their weight were distributed into the three weight categories as follows: 21 were *light*, 20 were *medium*, and 22 were *heavy*. For our binary classification problem, the users are 48.9% *light* and 51.1% *heavy*. While our lightest user was 115 lbs and our heaviest was 300 lbs, the majority of users fall between 120 and 180 lbs. Preliminary results indicated that it was extremely difficult to predict the underrepresented values at 275 and 300 pounds. In order to provide a more representative evaluation of our results, which are based on a relatively small data set to begin with, these two points were excluded from further analysis. Even with this, the results presented in Section 4.3 and displayed in Figure 3a show that the other set of outliers, between 220 lbs and 240 lbs, were also hard to predict. We should be able to do better in the future as we collect much more data. With these two data points removed, the average weight of the users is 160.5 lbs.

3.4 Induction Algorithms and Methodology

Our experiments involve three binary classification tasks, one for each trait, and two regression tasks, for height and weight prediction. Our models are all built using predictive data mining algorithms from the WEKA data mining suite [23]. We use WEKA’s instance based (IBk) “nearest neighbor” learner, multilayer perceptron neural network, and J48 decision tree algorithms for our

three classification tasks and use the instance based and neural network algorithms for our two regression tasks (J48 cannot predict numerical values). Default settings are used for all learning methods except for instance based, in which case we use 3 nearest neighbors ($k=3$) instead of the default of 1. Throughout this paper the algorithms are abbreviated as follows: IB3 for the instance based method, NN for the neural network method, and J48 for the decision tree method.

The experiment setup is very important. For this work it is critical that a user does not appear in both the training and test sets simultaneously, even if the training and test sets contain samples from different 10 second time periods. The reason that this is important is because we want to be able to identify characteristics of a user based on “universal” models generated from a set of other users. We have found in prior work that there is a tremendous improvement in performance if there is overlap for a user in the training and test sets—even if the overlap is for just a few examples from different time periods. To address this issue and to maximize the use of our very limited number of test subjects, we utilize leave-one-out cross validation, so that we train on all but one user and then test the induced model on that one held-out user; this process is repeated until all users are tested on. Thus, if we have n users in the data set, we will execute n runs in the leave-one-out cross validation process. All of our results are based on the aggregated values over these n runs.

Each test set will contain all of the examples for a single user, each one representing 10 seconds of walking activity. Most users have about 5 to 10 minutes of walking activity and the test set has, on average, 57 transformed 10-second examples. While we internally record our predictive performance on each of these 10-second examples, those results are not presented in this paper. Rather, for classification problems, our performance results are based on a simple majority-voting procedure. For example, if the test set has 57 examples for user *A* and 37 of those predict *male* and 20 predict *female*, we will classify the user as *male*. For regression tasks, rather than using a majority voting scheme, we aggregate the predictions by averaging the results (in the future we may consider a procedure that lessens the impact of outliers, like using the median value). The results using the aggregated predictions, as expected, perform better than the results based on individual examples, although the supporting data for this comparison is not presented in this paper.

4. RESULTS

In this section we present and evaluate the results associated with predicting the following three traits: sex, height, and weight. When it is not practical to show the detailed results for all of the data mining algorithms, we show it only for the instance based algorithm, since overall it tends to perform best.

4.1 Sex Classification Task Results

We begin with the results for predicting the sex of a user, because this binary classification problem is the simplest to evaluate. The confusion matrices for the sex classification task, for the instance based, neural network, and decision tree models, are provided in Table 3. As is the convention in data mining, the labels across the top row correspond to the predicted class values while the labels in the left-most column correspond to the actual class labels. Thus, Table 3a shows that of the 38 males in our data set, 31 are correctly classified as *male* and 7 are incorrectly classified as *female*. These confusion matrices are quite important because

they show us where the errors occur and allow us to make sure that the classifier is not trivial (i.e. always predicting the majority class).

Table 3. Confusion matrix for sex classification task

	Male	Female
Male	31	7
Female	12	16

(a) Instance Based (IB3)

	Male	Female
Male	30	8
Female	15	13

(b) Neural Network (NN)

	Male	Female
Male	27	11
Female	10	18

(c) Decision Tree (J48)

The accuracies associated with each of these models are summarized in Table 4. The MFC method corresponds to our “straw man” model that always predicts the Most Frequent Class. For results to be significant, they should outperform the MFC method, although with imbalanced data sets it might be possible for a model to be useful even if this is not the case, if it performs well on the rare class. In this case, our majority class contains 57.6% of the examples (i.e., 38 males out of a total of 66 subjects). Here the instance based method performs best (we denote the best method by underlining it), although its performance is far from ideal. We expect to improve our results in the future as we obtain additional training data and engineer more useful features.

Table 4. Accuracy for sex classification task

IB3	NN	J48	MFC
<u>71.2%</u>	65.2%	68.2%	57.6%

4.2 Height Prediction Task Results

Next we turn our attention to the numerical height trait. Because the native task here involves predicting a numerical value, our most detailed results are presented using a scatter plot. Due to space considerations, only the scatter plots for the instance based algorithm are provided. The results in Figure 2a are a traditional scatter plot, where each data point corresponds to a single user, whereas the same information is shown in Figure 2b, but the results for users with the same actual height are averaged together. The results in Figure 2a are of the most practical significance, because we care about performance on an individual basis, but the results in the more summarized Figure 2b are a bit easier to parse. In both cases the ideal performance would yield points on the line $y=x$, which corresponds to perfect predictions. A benchmark “straw man” strategy, analogous to the MFC strategy for classification tasks, is to always predict the average height, which is 69.1 inches. This straw man strategy is displayed in Figure 2 as a horizontal line. The line that optimally interpolates the *predicted* values is also shown and is labeled “linear fit.” Because the linear fit line is closer than the “straw man” line to the optimal strategy, represented by the line $y=x$, we conclude that our results clearly

outperform the straw man strategy of always predicting the average height.

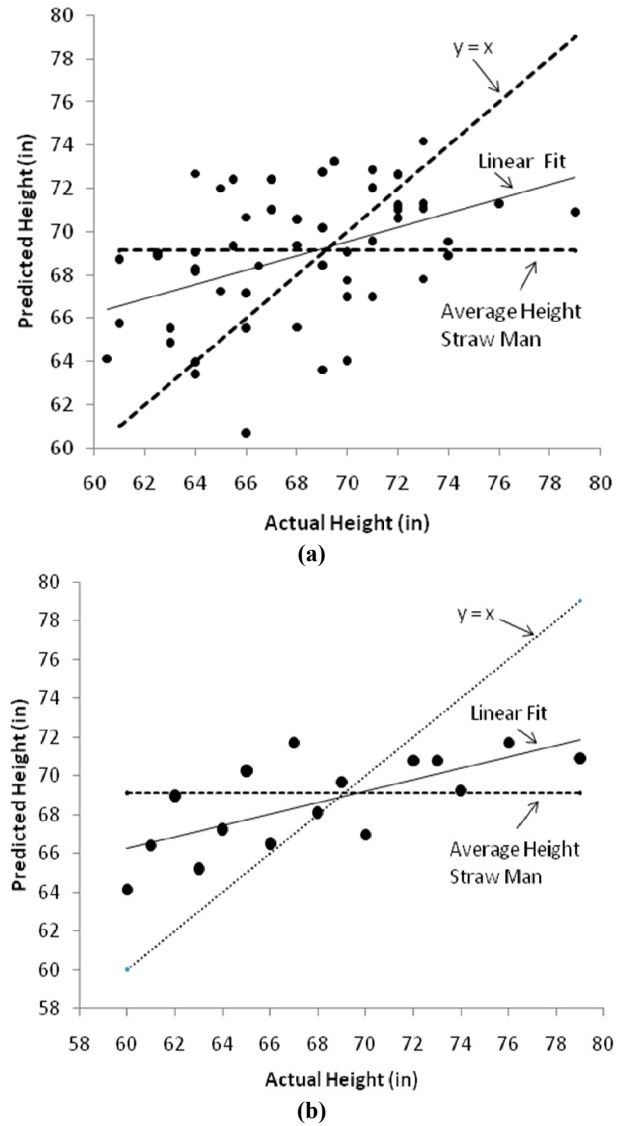


Figure 2. IB3 scatter plot results for predicting height. In (a) each point represents a single user and in (b) users with the same actual height are averaged.

Table 5 summarizes the performance of the numerical predictions using standard statistical measures (recall that J48 can only handle classification tasks). The metrics in Table 5 include “Mean Error”, which represents the mean absolute difference between the actual and predicted values for each user, and the Root Mean Squared Error (RMSE), which is computed by summing the square of the differences over all actual and predicted values and then taking the square root of that sum. For both metrics a smaller value indicates better performance. The main difference between these two metrics is that the RMSE assigns a higher penalty than Mean Error for large errors. RMSE can be smaller than Mean Error since values less than 1.0 get smaller when you square them.

Table 5. Summary statistics for numerical height prediction

Metric	IB3	NN	Straw
Mean Error	<u>3.23</u>	3.91	3.59
RMSE	<u>3.89</u>	3.92	4.17

The results in Table 5 show that IB3 performs best and outperforms the Straw Man strategy of always predicting the average height. But there is certainly room for future improvement.

As discussed earlier, we also evaluate our ability to classify a user as *short* or *tall*. This is a simple test to see if we can classify people in these two extremes. The confusion matrix results are shown in Table 6 and appear to be quite good.

Table 6. Confusion matrix results for height classification task

	Short	Tall
Short	15	5
Tall	2	20

(a) Instance Based (IB3)

	Short	Tall
Short	17	3
Tall	3	19

(b) Neural Network (NN)

	Short	Tall
Short	14	6
tall	3	19

(c) Decision Tree (J48)

The accuracies associated with classifying a user as *short* or *tall* are provided in Table 7. The results indicate that the model significantly outperforms the Most Frequent Class strategy and in fact performs much better than the model that predicts a user’s sex. In this case the neural network model performs best although the instance based model also performs quite well.

Table 7. Accuracy Results for height classification task

IB3	NN	J48	MFC
83.3%	<u>85.7%</u>	78.6%	52.4%

4.3 Weight Prediction Task Results

The instance based results for predicting the numerical weight trait, presented in the scatter plots in Figure 3, are presented in a manner similar to the results for height in Figure 2. However, because the range of weight values is so much greater than the range of height values, in Figure 3b users are aggregated into 10-pound bins. As we saw with height, the “linear fit” line that fits the predicted values comes much closer to the optimal performance (represented by the line $y=x$), than does the straw man strategy of always predicting the average weight of 160.5 lbs. But we do see some notable cases where our predicted weight is much higher than the actual observed weight. For example, in one case in Figure 3a a user weighing about 160 pounds is predicted to weigh about 240 pounds.¹ But overall the results are quite posi-

tive. For example, if we look at the three “actually” heaviest people in Figure 3a, we see that all are predicted to be substantially heavier than the average person. In general, we might expect that any model would have the most trouble predicting values at the extremes.

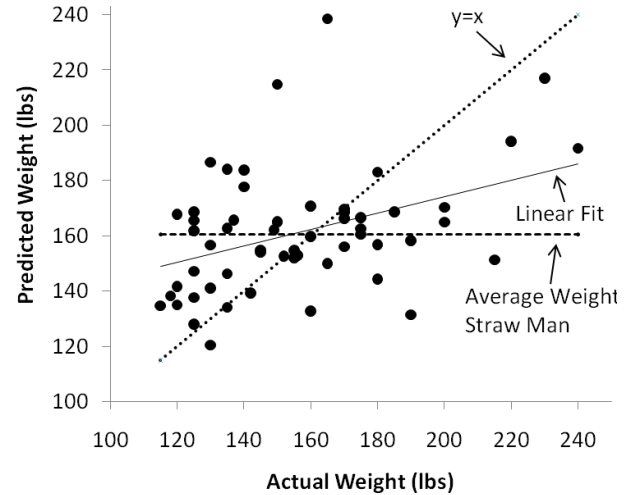
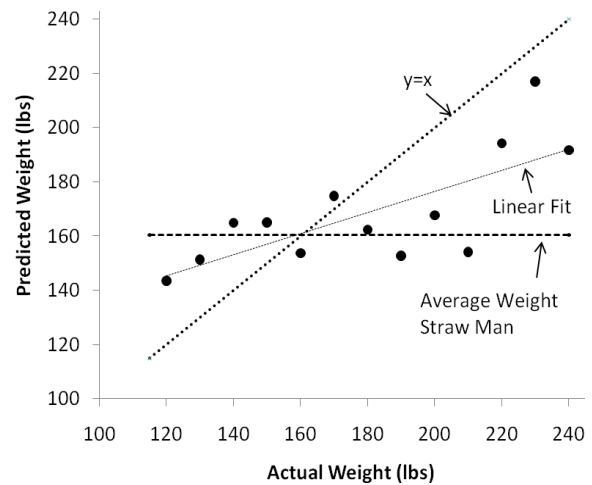
**(a)****(b)**

Figure 3. IB3 scatter plot results for predicting weight. In (a) each point represents a single user and in (b) users are grouped into 10-pound bins, then averaged.

The performance of the numerical weight predictions is summarized in Table 8. The results indicate that for Mean Error and Root Mean Squared Error, the instance based method performs best. It is likely that IB3 does poorly on the RMSE (i.e. not much better than the straw man) due to the few users that are predicted as being much heavier than their actual weights. The results indicate that, on average, the instance based algorithm produces predictions that are about 2 pounds more accurate than the Straw Man strategy of always predicting the average weight.

¹ There may be a good explanation for this error. Further analysis showed that this user’s questionnaire indicated that they had

knee problems that affected their ability to walk. One can easily imagine that this might make them appear heavier.

Table 8. Summary statistics for numerical weight prediction

Metric	IB3	NN	Straw
Mean Error	<u>22.61</u>	25.17	24.62
RMSE	<u>28.85</u>	32.33	29.52

The results in Table 9 show that we can do a relatively good job at distinguishing between *light* and *heavy* people, but that for the instance based and neural network models, we can do a much better job at identifying the *heavy* people.

Table 9. Confusion matrix for weight classification task

	Light	Heavy
Light	13	7
Heavy	2	17
(a) Instance Based (IB3)		
	Light	Heavy
Light	11	9
Heavy	5	14
(b) Neural Network (NN)		
	Light	Heavy
Light	15	5
Heavy	5	14
(c) Decision Tree (J48)		

The accuracy results associated with classifying a user as being light or heavy are summarized in Table 10. The results indicate that the instance based method performs best and that we can significantly outperform the Most Frequent Class strategy. But there is substantial room for future improvement.

Table 10. Accuracy Results for weight classification task

IB3	NN	J48	MFC
<u>78.9%</u>	65.0%	76.3%	51.3%

5. RELATED WORK

The use of patterns of movement for measuring “soft” biometric traits, such as gait, height, weight, and sex, is a relatively new but growing area. Most of the work in this area is geared toward biometric identification, where research has demonstrated that measuring these soft traits can improve the performance of biometric identification systems [2,9,14]. But the connection to work in biometrics is not critical, since information about these “soft” traits can be used for other purposes—which is part of our motivation for this work. Furthermore, as will be made clear in this section, the measurement and impact of these soft traits on activities such as walking has been studied for many years in other disciplines.

Our prior work on biometric identification [10] is highly relevant to the work in this paper. That work used the same WISDM platform [12,21] that is used in this paper and also was based on accelerometer data from Android smart phones. The main difference is that in the prior work the goal was to identify users whereas the goal here is to learn about their traits. But of course the identification was only possible because the predictive models *implicitly* identified user traits related to how they move. In that prior work we demonstrated that a smart phone’s accelerometer was suffi-

cient for identifying a user from a pool of 36 users with 100% accuracy, given only on a few minutes of accelerometer data (some of which was available for training). The data that we supplied was from a variety of physical activities whereas in this work we simplify things by only using *walking* data. There have also been numerous other accelerometer-based biometric identification studies, where accelerometers were placed on one more body parts [6,7,13]. But it should be noted that there other means for using motion to perform biometric identification, most notably vision-based systems [17], which tend to be more popular than accelerometer-based systems.

Some of this work on accelerometer based biometric identification, which relies primarily on gait, mentions factors that impact the results. Because these factors impact gait, it seems reasonable to conclude that they *may* be predictable based on gait and are thus candidates as “soft” traits. One study found that the weight of one’s shoes can impact gait-based recognition performance [7]. A second study found that an imposter could not mimic a person’s gait to fool a gait-based recognition system unless they also knew the person’s sex [7]. Finally, a third study found that biometric identification can be improved by estimating a person’s height and stride length [4].

Other research communities, including those that study ergonomics and kinesiology, have also studied factors and traits that impact gait. One such study showed that the texture of footwear impacts gait [18]. Other footwear related studies showed that the type of shoe impacts gait [20] and that heel height impacts gait patterns in women [16]. Not coincidentally, our survey tracks information on such things as shoe type—and even before finding the research on heel height one of our female researchers agreed to provide walking data while wearing high heels. Another interesting study examined the interaction between gait speed, obesity, and race/ethnicity and found observable patterns [24]. The study found that gait speed is impacted by obesity, but that this impact is substantially affected by the subject’s race/ethnicity. The connection between gait speed and obesity supports our belief that weight can be estimated using accelerometer data collected during walking (some of our features capture gait speed) and we do track race in our questionnaire. Other research showed that gait does in fact vary between men and women [5] while a related study examined gender recognition based on movement (using point light displays to hide other cues) and demonstrated that viewers believe that males move their shoulders more and women move their hips more [3]. Note that our experiments, which involve accelerometer readings based on smart phones located in ones pants pocket, should be able to detect differences in hip movements, if they exist. A final study showed that accelerometer data while standing can tell us something about a person’s balance [15].

Based on these studies, we see that there are reasons to believe that we might be able to identify many soft traits that are associated with a user, including some, like balance, that are not listed in Table 1. But we also see that we may be able to identify other, more temporary things about a user, such as what type of shoe they are currently wearing. But there is also the possibility for learning about a person based on longer term trends, whereas our research thus far, and most of the research discussed in this section, is about short term observations. For example, adolescents have different activity patterns than others and in fact these patterns differ by gender [8]. Also, a longer term study showed that an accelerometer can estimate walking speed together with the pattern, intensity and duration of daily walking activity [19].

Thus, this mining of longer term sensor data is a potentially very profitable area for research. In fact, the WISDM project is well positioned for this type of research since our ActiTracker [1] cell phone application, based on our activity recognition research [10], will allow us to track activity information for users over long periods of time—and provide the information back to them via a web interface. This service is under active development and should be released to the Android marketplace in the near future.

6. CONCLUSIONS AND FUTURE WORK

The use of smart phone sensors to learn about users is an interesting and timely topic, and smart phone-based sensor mining application should proliferate over the next decade. Many applications will undoubtedly be developed that cannot easily be imagined at the present time. In this paper we take a step toward the use of smart phone sensors to identify and measure user traits. We show that it is possible to identify a user's sex, whether they are short or tall, and whether they are light or heavy, although the performance of these predictions is not consistently good (but they are better than always guessing the most frequent class). Similarly, we demonstrate that we can predict a user's height and weight, but that such fine grained predictions are of only modest quality. In the immediate future we will work to improve these results. Certainly additional training data will help, since we currently have useful data from less than 70 users. Additional work will also go into generating more informative high level features, including some that better capture the time-series nature of the data. We also will begin to investigate the use of other sensors on the smart phone, especially the GPS sensor.

In addition to improving current results, we plan to predict the values of many additional traits. We will begin with the traits listed in Table 1, but also consider other traits—including traits that may only become apparent based on long term patterns. Finally, we plan to implement some of our predictive models so that users can get results in real-time, via our WISDM platform. We may do this with a research-oriented app or by building some of these capabilities into the ActiTracker [1] application/service that we are developing. We believe that we will learn much more about this new and exciting research area by actually building and deploying publicly available sensor mining applications.

Finally, there are other issues that need to be addressed, such as privacy and security. We address this topic in some detail in a recent paper describing our WISDM sensor mining architecture and platform [12]. But there also needs to be some education of the public about the benefits and dangers of this technology, since no application can be made perfectly secure and no data can always be guaranteed to remain private. But many innovative and creative smart phone-based sensor mining applications are just waiting to be developed and we hope that this paper will advance research in this area as well as bring attention to the topic.

7. REFERENCES

- [1] ActiTracker. <http://ActiTracker.com/>
- [2] Ailisto, H., Vildjounaite, E., Lindholm, M., Mäkelä, S., and Peltola, J. 2006. Soft biometrics—combining body weight and fat measurements with fingerprint biometrics, *Pattern Recognition Letters*, 27(5): 325-334.
- [3] Barclay, C. Cutting, J. and Kozlowski, L. 1978. Temporal and spatial factors in gait perception that influence gender recognition, *Perception and Psychophysics*, 23:145–152.
- [4] BenAbdelkader, C., Cutler, R., and Davis, L.S. 2002. Person identification using automatic height and stride estimation. In *Proceedings of the 16th International Conference on Pattern Recognition*, vol. 4, 377-380.
- [5] Cho S.H., Park J.M., and Kwon O.Y. 2004. Gender differences in three dimensional gait analysis data from 98 healthy Korean adults. *Clinical Biomechanics*, 19:145-152.
- [6] Gafurov, D., Helkala, K., and Sondrol, T. 2006. Biometric gait authentication using accelerometer sensor, *Journal of Computers*, 1 (7): 51-59.
- [7] Gafurov, D. and Snekenes, E. 2008. Gait recognition using wearable motion recording sensors, In *EURASIP Journal on Advances in Signal Processing*.
- [8] Jago R, Anderson C.B., Baranowski T, and Watson K. 2005. Adolescent patterns of physical activity: differences by gender, day, and time of day. *American Journal of Preventative Medicine*, 28:447–452.
- [9] Jain, A. K., Dass, S. C., and Nandakumar, K. 2004. Can soft biometric traits assist user recognition? In *Proceedings of SPIE International Symposium on Defense and Security: Biometric Technology for Human Identification*, Vol. 5404, 561-572, Orlando, FL.
- [10] Kwapisz, J.R., Weiss, G. M., and Moore, S.A. 2010. Activity recognition using cell phone accelerometers. In *Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data*, Washington DC, 10-18.
- [11] Kwapisz, J.R., Weiss, G.M., and Moore, S.A. 2010. Cell phone-based biometric identification. In *Proceedings of the IEEE 4th International Conference on Biometrics: Theory, Applications and Systems (BTAS-10)*, Washington DC.
- [12] Lockhart, J.W. and Weiss, G. 2011. Design considerations for the WISDM smart phone-based sensor mining architecture. In *Proceedings of the Fifth International Workshop on Knowledge Discovery from Sensor Data*, San Diego, CA.
- [13] Mantyjarvi, J., Lindholdm, M., Vildjounaite, E., Makela, S., and Ailisto, H. 2005. Identifying users of portable devices from gait pattern with accelerometers. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 973-976.
- [14] Marcialis, G.L., Roli, F., and Muntoni, D. 2009. Group-specific face verification using soft biometrics, *Journal of Visual Languages and Computing*, 20(2): 101–109.
- [15] Mayagoitia, R.E., Lotters, J.C., Veltink, P.H., and Hermens, H. 2002. Standing balance evaluation using a triaxial accelerometer, *Gait & Posture*, 16 (1): 55-59. DOI: 10.1016/S0966-6362(01)00199-0.
- [16] Merrifield, H.H. 1971. Female gait patterns in shoes with different heel heights, *Ergonomics*, 14(3): 411–417.
- [17] Nixon, M., Tan, T. and Chellappa, R. 2006. *Human Identification Based on Gait*. New York: Springer Science + Business Media Inc., Chapter. 1.
- [18] Nurse, M.A., Hulliger, M., Wakeling, J.M., Nigg, B.M., and Stefanyshyn, D.J. 2005. Changing the texture of footwear can alter gait patterns, *Journal of Electromyography and Kinesiology*, 15(5): 496-506. DOI: 10.1016/j.jelekin.2004.12.003

- [19] Schutz, Y., Weinsier, S., Terrier, P., and Durrer, D. 2002. A new accelerometric method to assess the daily walking practice. *International Journal of Obesity & Related Metabolic Disorders*, 26(1):111-118.
- [20] Soames, R. W. and Evans, A.A. 1987. Female gait patterns: The influence of footwear, *Ergonomics*, 30(6): 893-900.
- [21] Weiss, G. M. 2011. WISDM (Wireless Sensor Data Mining) Project. Fordham University, Department of Computer and Info. Science, <http://www.cis.fordham.edu/wisdm/>
- [22] Weiss, G. M., and Hirsh, H. 1998. Learning to predict rare events in event sequences, In *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, AAAI Press, Menlo Park, CA, 359-363.
- [23] Witten, I. H. and Frank, E. 2005. *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. Morgan Kaufmann.
- [24] Xu B., Houston D.K., Gropper S.S., and Zizza C.A. 2009. Race/Ethnicity differences in the relationship between obesity and gait speed among older Americans, *Journal of Nutrition for the Elderly*, Oct;28(4):372-85.